

Документ подписан простой электронной подписью
Информация о владельце:
ФИО: Гаранин Максим Александрович
Должность: Ректор
Дата подписания: 19.06.2025 12:44:59
Уникальный программный ключ:
7708e3a47e66a8ee02711b298d7c78bd1e40bf88

ФЕДЕРАЛЬНОЕ АГЕНТСТВО ЖЕЛЕЗНОДОРОЖНОГО ТРАНСПОРТА
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«ПРИВОЛЖСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ПУТЕЙ СООБЩЕНИЯ»

ОБЩЕПРОФЕССИОНАЛЬНЫЙ МОДУЛЬ
"СИСТЕМЫ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА"
Практикум по машинному обучению
рабочая программа дисциплины (модуля)

Специальность Специальность 23.05.06 Строительство железных дорог, мостов и транспортных тоннелей
Специализация Управление техническим состоянием железнодорожного пути

Квалификация **Инженер путей сообщения**

Форма обучения **заочная**

Общая трудоемкость **2 ЗЕТ**

Виды контроля на курсах:
зачеты 5

Распределение часов дисциплины по курсам

Курс	5		Итого	
	уп	ип		
Лабораторные	8	8	8	8
Конт. ч. на аттест. в период ЭС	0,15	0,15	0,15	0,15
Итого ауд.	8	8	8	8
Контактная работа	8,15	8,15	8,15	8,15
Сам. работа	60	60	60	60
Часы на контроль	3,85	3,85	3,85	3,85
Итого	72	72	72	72

Программу составил(и):

к.п.н., доцент, Тюжина И.В.

Рабочая программа дисциплины

Практикум по машинному обучению

разработана в соответствии с ФГОС ВО:

Федеральный государственный образовательный стандарт высшего образования - специалитет по специальности 23.05.06 Строительство железных дорог, мостов и транспортных тоннелей (приказ Минобрнауки России от 27.03.2018 г. № 218)

составлена на основании учебного плана: 23.05.06-25-1-СЖДп.plz.plx

Специальность Специальность 23.05.06 Строительство железных дорог, мостов и транспортных тоннелей Направленность (профиль) Управление техническим состоянием железнодорожного пути

Рабочая программа одобрена на заседании кафедры

Цифровые технологии

Зав. кафедрой

1. ЦЕЛИ ОСВОЕНИЯ ДИСЦИПЛИНЫ (МОДУЛЯ)

1.1	Применение на практике методов машинного обучения при решении задач профессиональной деятельности.
-----	--

2. МЕСТО ДИСЦИПЛИНЫ (МОДУЛЯ) В СТРУКТУРЕ ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ

Цикл (раздел) ОП:	Б1.О.42.02
-------------------	------------

3. КОМПЕТЕНЦИИ ОБУЧАЮЩЕГОСЯ, ФОРМИРУЕМЫЕ В РЕЗУЛЬТАТЕ ОСВОЕНИЯ ДИСЦИПЛИНЫ (МОДУЛЯ)

ОПК-10 Способен формулировать и решать научно-технические задачи в области своей профессиональной деятельности

ОПК-10.3 Решает задачи в области профессиональной деятельности, используя перспективные методы машинного обучения

В результате освоения дисциплины (модуля) обучающийся должен

3.1	Знать:
3.1.1	- основные классы научно-технических задач в области своей профессиональной деятельности, решаемые методами машинного обучения;
3.1.2	- основные классы интеллектуальных информационных систем, необходимые в процессе решения научно-технических задач в области своей профессиональной деятельности;
3.2	Уметь:
3.2.1	- использовать методы искусственного интеллекта для решения прикладных задач;
3.3	Владеть:
3.3.1	- навыками решения научно-технических задач в области своей профессиональной деятельности с использованием технологий анализа данных и машинного обучения.

4. СТРУКТУРА И СОДЕРЖАНИЕ ДИСЦИПЛИНЫ (МОДУЛЯ)

Код занятия	Наименование разделов и тем /вид занятия/	Семестр / Курс	Часов	Примечание
	Раздел 1. Практикум по машинному обучению			
1.1	Подготовка и фильтрация данных. Распространенные форматы данных: csv, json. Удаление поврежденных значений. Выбор данных по заданному признаку и группе признаков. /Лаб/	5	2	
1.2	Преобразование данных Формирование нового столбца данных на основе существующих. Формирование вспомогательного столбца данных. Изменение типа данных столбца. /Ср/	5	4	
1.3	Агрегация данных Группировка данных по признаку. Агрегация данных: подсчет среднего, общего, максимального и минимального значений по определенному признаку. Выбор оптимального метода. /Ср/	5	4	
1.4	Определение закономерностей в опозданиях железнодорожного транспорта. Графическая обработка данных. Построение графиков зависимости. /Лаб/	5	2	
1.5	Определение закономерностей в опоздании железнодорожного транспорта. Построение графиков распределения точек, определение числа кластеров, формируемых наблюдениями. Построение тепловой карты. Различные виды графиков (seaborn, matplotlib, встроенная визуализация pandas). /Ср/	5	2	
1.6	Определение неисправности железнодорожного полотна. Обучение модели. Подготовка данных. Разделение данных на тренировочный, тестовый и валидационные датасеты. Построение нейронной сети. /Лаб/	5	4	
1.7	Исследование данных о дорожных происшествиях и серьезности аварий. Дерево решений. Построение матрицы ошибок (confusion matrix) Подсчет метрик качества (Precision, Recall, F1 score). /Ср/	5	4	
1.8	Исследование данных о дорожных происшествиях и серьезности аварий. Случайный лес Выбор оптимальных параметров дерева решений. Создание модели с указанным числом деревьев и максимальной глубины. Предсказание класса по тренировочной выборке. Отбор топ-5 самых важных переменных для классификации /Ср/	5	4	

1.9	Определение тональности текста. Нейронные сети. Определение тональности текста с помощью нейронных сетей. Библиотеки tensorflow, Keras, набор данных IMDB. /Ср/	5	6	
	Раздел 2. Самостоятельная работа			
2.1	Подготовка и фильтрация данных. Регистрация на портале kaggle. Распространенные форматы данных: csv, json. Подготовка к лабораторным работам. /Ср/	5	2	
2.2	Преобразование данных. Методы работы структурами данных pandas. Операции с pandas. series. Повторение материала, изученного в курсе «Технологии искусственного интеллекта». Подготовка к лабораторным работам. /Ср/	5	4	
2.3	Агрегация данных. Группировка. Функции агрегирования в pandas. Повторение материала, изученного в курсе «Технологии искусственного интеллекта». Подготовка к лабораторным работам. /Ср/	5	4	
2.4	Определение закономерностей в опозданиях железнодорожного транспорта. Графическая обработка данных. Методы визуализации pandas. Библиотеки matplotlib и seaborn. Параметры методов. Повторение материала, изученного в курсе «Технологии искусственного интеллекта». Подготовка к лабораторным работам. ЭИОС СамГУПС 4 /Ср/	5	4	
2.5	Определение неисправности железнодорожного полотна. Обучение модели. Работа с изображениями в python. Построение нейронной сети. Виды слоев. Повторение теоретического материала, изученного в курсе «Технологии искусственного интеллекта». Подготовка к лабораторным работам. /Ср/	5	4	
2.6	Определение закономерностей в опозданиях железнодорожного транспорта. Графическая обработка данных. Графическая обработка данных. Методы визуализации pandas. Библиотеки matplotlib и seaborn. Параметры методов. Повторение материала, изученного в курсе «Технологии искусственного интеллекта». Подготовка к лабораторным работам. /Ср/	5	4	
2.7	Исследование данных о дорожных происшествиях и серьезности аварий. Случайный лес. Основные параметры случайного леса. Повторение теоретического материала, изученного в курсе «Технологии искусственного интеллекта». Подготовка к лабораторным работам. /Ср/	5	4	
2.8	Определение тональности текста. Нейронные сети Методы библиотеки tensorflow. Подготовка к лабораторным работам. /Ср/	5	6	
2.9	Набор данных IMDB. Исследование данных. Библиотека keras. Изучение синтаксиса методов библиотеки. Подготовка к лабораторной работе. /Ср/	5	4	
	Раздел 3. Контактные часы на аттестацию			
3.1	Зачет /КЭ/	5	0,15	

5. ОЦЕНОЧНЫЕ МАТЕРИАЛЫ

Оценочные материалы для проведения промежуточной аттестации обучающихся приведены в приложении к рабочей программе дисциплины.

Формы и виды текущего контроля по дисциплине (модулю), виды заданий, критерии их оценивания, распределение баллов по видам текущего контроля разрабатываются преподавателем дисциплины с учетом ее специфики и доводятся до сведения обучающихся на первом учебном занятии.

Текущий контроль успеваемости осуществляется преподавателем дисциплины (модуля) в рамках контактной работы и самостоятельной работы обучающихся. Для фиксирования результатов текущего контроля может использоваться ЭИОС.

6. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ (МОДУЛЯ)				
6.1. Рекомендуемая литература				
6.1.1. Основная литература				
	Авторы, составители	Заглавие	Издательство, год	Эл. адрес
Л1.1	Никольский С. Н.	Автоматизация информационного поведения и искусственный интеллект: Учебное пособие	Москва: МИРЭА, 2020	https://e.lanbook.com/bc
6.2 Информационные технологии, используемые при осуществлении образовательного процесса по дисциплине (модулю)				
6.2.1 Перечень лицензионного и свободно распространяемого программного обеспечения				
6.2.2 Перечень профессиональных баз данных и информационных справочных систем				
7. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ (МОДУЛЯ)				

**ОЦЕНОЧНЫЕ МАТЕРИАЛЫ ДЛЯ ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ
ПО ДИСЦИПЛИНЕ (МОДУЛЮ)**

ПРАКТИКУМ ПО МАШИННОМУ ОБУЧЕНИЮ

(наименование дисциплины(модуля))

Направление подготовки / специальность

**23.05.06 Строительство железных дорог, мостов и транспортных
тоннелей**

(код и наименование)

Направленность (профиль)/специализация

Управление техническим состоянием железнодорожного пути
(наименование)

Содержание

1. Пояснительная записка.
2. Типовые контрольные задания или иные материалы для оценки знаний, умений, навыков и (или) опыта деятельности, характеризующих уровень сформированности компетенций.
3. Методические материалы, определяющие процедуру и критерии оценивания сформированности компетенций при проведении промежуточной аттестации.

1. Пояснительная записка

Цель промежуточной аттестации – оценивание промежуточных и окончательных результатов обучения по дисциплине, обеспечивающих достижение планируемых результатов освоения образовательной программы.

Формы промежуточной аттестации: Зачёт, 8 семестр/ЗФО 5 курс

Перечень компетенций, формируемых в процессе освоения дисциплины

Код и наименование компетенции	Код индикатора достижения компетенции
ОПК-10: Способен формулировать и решать научно-технические задачи в области своей профессиональной деятельности	ОПК-10.3: Решает задачи в области профессиональной деятельности, используя перспективные методы машинного обучения

Результаты обучения по дисциплине, соотнесенные с планируемыми результатами освоения образовательной программы

Код и наименование индикатора достижения компетенции	Результаты обучения по дисциплине	Оценочные материалы
ОПК-10.3: Решает задачи в области профессиональной деятельности, используя перспективные методы машинного обучения	Обучающийся знает : - основные классы научно-технических задач в области своей профессиональной деятельности, решаемые методами машинного обучения; - основные классы интеллектуальных информационных систем, необходимые в процессе решения научно-технических задач в области своей профессиональной деятельности;	Вопросы (№1 - №30)
	Обучающийся умеет : - использовать методы искусственного интеллекта для решения прикладных задач;	Задания (№1 - №30)
	Обучающийся владеет : - навыками решения научно-технических задач в области своей профессиональной деятельности с использованием технологий анализа данных и машинного обучения.	Задания (№ 36- №40)

Промежуточная аттестация (зачет) проводится в одной из следующих форм:

- 1) собеседование;
- 2) выполнение заданий в ЭИОС университета.

2. Типовые¹ контрольные задания или иные материалы для оценки знаний, умений, навыков и (или) опыта деятельности, характеризующих уровень сформированности компетенций

2.1. Типовые вопросы (тестовые задания) для оценки знаниевого образовательного результата

Проверяемый образовательный результат:

Код и наименование индикатора достижения компетенции	Образовательный результат
	Обучающийся знает :

¹ Приводятся типовые вопросы и задания. Оценочные средства, предназначенные для проведения аттестационного мероприятия, хранятся на кафедре в достаточном для проведения оценочных процедур количестве вариантов. Оценочные средства подлежат актуализации с учетом развития науки, образования, культуры, экономики, техники, технологий и социальной сферы. Ответственность за нераспространение содержания оценочных средств среди обучающихся университета несут заведующий кафедрой и преподаватель – разработчик оценочных средств.

ОПК-10.3: Решает задачи в области профессиональной деятельности, используя перспективные методы машинного обучения	- основные классы научно-технических задач в области своей профессиональной деятельности, решаемые методами машинного обучения; - основные классы интеллектуальных информационных систем, необходимые в процессе решения научно-технических задач в области своей профессиональной деятельности;
Задание 1 Машинное обучение — a) специализированное программное решение (или набор решений), которое включает в себя все инструменты для извлечения закономерностей из сырых данных b) эта группировка объектов (Наблюдений, событий) на основе данных, описывающих свойства объектов c) набор данных, каждая запись которого представляет собой учебный пример, содержащий заданный входной набор данных, и соответствующий ему правильный выходной результат d) подразделение искусственного интеллекта изучающий методы построения алгоритмов, способных обучаться на данных Задание 2 Среди предложенных задач машинного обучения укажите задачи регрессии a) Поиск негативных отзывов на фильм на сайте кинокомпании b) Алгоритм фильтрации спама c) Предсказание срока окупаемости проекта d) Предсказание рыночной стоимости квартиры e) Поиск мошеннических транзакций Задание 3 К библиотекам анализа данных в Python относятся ... a) Tkinter b) Pandas c) NumPy d) Matplotlib Задание 4 Какие из перечисленных команд вернут данные для России и Зимбабве? a) <code>res.query('country == ["Russia","Zimbabwe"]')</code> b) <code>res.query('country == "Russia" "Zimbabwe"')</code> c) <code>res.query('country == "Russia" country == "Zimbabwe"')</code> d) <code>res.query('country == "Russia" & country == "Zimbabwe"')</code> e) <code>res[res.country == "Russia" res.country == "Zimbabwe"]</code> f) <code>res[(res.country == "Russia") (res.country == "Zimbabwe")]</code> Задание 5 Отметьте верные утверждения об алгоритме случайного леса - Random Forest a) Итоговым предсказанием модели является предсказание случайного дерева b) Параметры для каждого дерева (глубина, минимальное число образцов в листе и т.д.) выбираются случайно c) Каждое дерево в лесу получает случайный поднабор данных d) Число деревьев в лесу выбирается случайным образом e) Предсказание модели - усреднённые предсказания деревьев Задание 6 Какой атрибут отвечает за минимальное число образцов в листе? a) <code>min_samples_leaf</code> b) <code>max_depth</code> c) <code>min_samples_split</code> d) <code>min_impurity_decrease</code> Задание 7 Какой атрибут отвечает за минимальное число образцов в узле для разделения? a) <code>min_samples_leaf</code> b) <code>max_depth</code> c) <code>min_samples_split</code> d) <code>min_impurity_decrease</code> Задание 8 Отметьте верные утверждения о влиянии параметров решающего дерева на переобучение a) Чем меньше значение <code>min_samples_split</code>, тем меньше тенденция к переобучению b) Чем меньше значение <code>min_samples_leaf</code>, тем меньше тенденция к переобучению c) Чем меньше глубина дерева, тем меньше тенденция к переобучению d) Глубина дерева никак не влияет на его переобученность	

- e) Чем меньше значение параметра `max_leaf_nodes`, тем меньше тенденция к переобучению

Задание 9

Какой классификатор необходимо использовать для предсказания размера одобренного кредита по заданным параметрам?

- a) `DecisionTreeRegressor`
- b) `DecisionTreeClassifier`
- c) Может быть использован и `DecisionTreeClassifier`, и `DecisionTreeRegressor`

Задание 10

Отметьте верные утверждения о деревьях решений.

- a) чем меньше свойство `gini`, тем однороднее примеры в листе
- b) свойство `gini` определяется только для листьев, но не для узлов
- c) свойство `samples` указывает на количество примеров в узле
- d) параметр `gini`, во всех листьях должен иметь одинаковые значения

Задание 11

Отметьте верные утверждения о случайном лесе

- a) чем больше деревьев в лесу, тем лучше качество предсказания
- b) время настройки и работы случайного леса увеличивается пропорционально количеству деревьев в лесу.
- c) случайный лес применим для решения задач регрессии
- d) случайный лес применим для решения задач классификации

Задание 12

Отметьте верные утверждения

- a) `RandomizedSearchCV` находит лучшую комбинацию параметров для дерева решений
- b) `RandomizedSearchCV` работает быстрее, чем `GridSearchCV`
- c) `GridSearchCV` находит лучшую комбинацию параметров для дерева, из тех значений параметров, что были переданы
- d) `RandomizedSearchCV` дает более точное предсказание, чем `GridSearchCV`

Задание 13

Выберите из списка задачи классификации:

- a) разделить по фотографии животных на кошек и собак
- b) предсказать стоимость квартиры по заданным параметрам
- c) разделить грибы на съедобные и ядовитые
- d) сгруппировать тексты по их эмоциональной окраске

Задание 14

Отметьте верные утверждения

- a) В качестве переменных для обучения дерева могут быть использованы как количественные, так и категориальные признаки
- b) Деревья решений могут решать задачу классификации с произвольным числом классов
- c) Чем меньше значение `min_samples_leaf` тем точнее будут предсказания на тренировочной выборке
- d) Чем меньше значение `min_samples_leaf` тем точнее будут предсказания на тестовой выборке

Задание 15

Обучающая выборка (`X_text`, `y_test`) это –

- a) Выборка, по которой настраиваются оптимальные параметры дерева
- b) Выборка, по которой оценивается качество полученного дерева решений
- c) Выборка, по которой осуществляется выбор наилучшей модели из множества моделей
- d) Множество целевых значений для данного набора данных

Задание 16

При обучении некоторой модели на тренировочной выборке и оценке её качества на тестовой выборке, получена большая разница между значениями метрик на тренировочной и тестовой выборках. О чем это может говорить?

- a) Модель переобучена
- b) Распределение данных в тестовой выборке сильно отличается от распределения в тренировочной
- c) Модель недообучена
- d) Модель идеально настроена

Задание 17

`Data` – `pandas.DataFrame`, размером 20 на 20. Какие из предложенных комбинаций не вызовут ошибку?

- a) `Data.iloc[[1,4,5],0:3]`
- b) `Data.iloc[[0,15], [1,5,-1]]`
- c) `Data.iloc[[1,4:15],0:3]`
- d) `Data[1:5,[1,6]]`

Задание 18

Какие строки позволяют отобразить 5 первых строк датафрейма data (индексы последовательность чисел от 0 с шагом 1)

- a) data.iloc[:5]
- b) data.head(5)
- c) data.loc[:5]
- d) data.loc[:4]
- e) data.tail(5)

Задание 19

В каких случаях может быть использована ступенчатая функция активации?

- a) Определить к съедобным или несъедобным относится гриб?
- b) Определить к какому из трех классов (setosa, versicolor, virginica) относится ирис?
- c) Разделить тексты на нейтральные, негативные и положительные.
- d) Предсказать размер заработной платы специалиста по определенным параметрам.
- e) Разделить изображения на содержащие и не содержащие открытый огонь.

Задание 20

Какое значение может иметь взвешенная сумма входов нейрона (до использования активационной функции)

- a) [-1,1]
- b) [0,1]
- c) (-1,1)
- d) $(-\infty, +\infty)$

Оценочный лист к типовому заданию А (модельный ответ):

A1	A2	A3	A4	A5	A6	A7	A8	A9	A10
b	d	b	b	b	a	a,c	d	d	b
A11	A12	A13	A14	A15	A16	A17	A18	A19	A20
b,c,d	b,c	b	b,d	b	a	a	d	b	a,b,c
A21	A22	A23	A24	A25	A26	A27	A28	A29	A30
c	a	c	b	d	a	b	a	d	b

2.2. Типовые задания для оценки навыкового образовательного результата

Проверяемый образовательный результат:

Код и наименование индикатора достижения компетенции	Образовательный результат
ОПК-10.3: Решает задачи в области профессиональной деятельности, используя перспективные методы машинного обучения	Обучающийся умеет: - использовать методы искусственного интеллекта для решения прикладных задач;

Приложение 2

Задание №1

Для ответа на этот вопрос вам понадобится набор данных imdb из библиотеки keras from keras.datasets
import imdb

Определите, какое слово стоит на 1000 месте (имеет индекс 1000) в частотном словаре датасета IMDb?

Запишите ответ:

1)	Ответ:	
----	--------	--

Задание №2

Для ответа на этот вопрос вам понадобится датасет imdb из библиотеки keras `from keras.datasets import imdb`
Определите, какое слово стоит на 3355 месте (имеет индекс 3355) в частотном словаре датасета IMDb?

Запишите ответ:

1)	Ответ:	
----	--------	--

Задание №3

Для ответа на этот вопрос вам понадобится датасет imdb из библиотеки keras `from keras.datasets import imdb`

Определите на каком месте в списке самых часто встречающихся слов в наборе данных IMDb стоит слово "joker" (впишите полученный индекс)?

Запишите ответ:

1)	Ответ:	
----	--------	--

Задание №4

Для ответа на этот вопрос вам понадобится датасет imdb из библиотеки keras
`from keras.datasets import imdb`

Определите на каком месте в списке самых часто встречающихся слов в наборе данных IMDb стоит слово "study" (впишите полученный индекс)?

Запишите ответ:

1)	Ответ:	
----	--------	--

Задание №5

В какой стране значение соотношения между количеством вакцинированных и общей численностью населения на текущую дату в стране максимально? Поле Daily vaccinations per million.

Источник данных - country_vaccinations.csv

Запишите ответ:

1)	Ответ:	
----	--------	--

Задание №6

В какой строке (укажите индекс) находится максимальное значение по полю `daily_vaccinations_per_million` для стран Италия, Испания и Португалия.

Ответ - одно число, индекс строки.

Источник данных - `country_vaccinations.csv`

Запишите число:

1)	Ответ:	
----	--------	--

Задание №7

В какой строке указаны максимальные показатели по `daily_vaccinations_per_million` для Аргентины, России и Уэльса.

Укажите номера строк через пробел. Например: 45 678 2223

Источник данных - `country_vaccinations.csv`

Запишите ответ:

1)	Ответ:	
----	--------	--

Задание №8

Укажите название библиотеки позволяющей строить диаграммы рассеивания, графики и гистограммы в Python

Запишите ответ:

1)	Ответ:	
----	--------	--

Задание №9

Имеется набор данных о работе датчиков в грузовых вагонах поезда. Необходимо по датчикам автоматически разделять вагоны на полностью заполненные и пустые.

Напишите команду, создающую дерево решений `dt` для этого набора.

Запишите ответ:

1)	Ответ:	
----	--------	--

Задание №10

Какое значение имеет параметр daily_vaccinations для строки с индексом 3567?

Источник данных - country_vaccinations.csv

Запишите ответ:

1)	Ответ:	
----	--------	--

Задание №11

Какое значение имеет параметр daily_vaccinations строки датасета country_vaccinations с индексом 3367?

Источник данных - country_vaccinations.csv

Запишите ответ:

1)	Ответ:	
----	--------	--

Задание №12

Какое значение имеет параметр daily_vaccinations строки с индексом 25?

Источник данных - country_vaccinations.csv

Запишите ответ:

1)	Ответ:	
----	--------	--

Задание №13

Какое значение имеет параметр people_fully_vaccinated_per_hundred для строки датасета с индексом 4036?

Источник данных - country_vaccinations.csv

Запишите ответ:

1)	Ответ:	
----	--------	--

Задание №14

Какое значение имеет параметр `people_vaccinated` для последней строки датасета `country_vaccinations`?

Источник данных - `country_vaccinations.csv`

Запишите ответ:

1)	Ответ:	
----	--------	--

Задание №15

Какое значение имеет параметр `total_vaccinations_per_hundred` для строки датасета с индексом 4036?

Источник данных - `country_vaccinations.csv`

Запишите ответ:

1)	Ответ:	
----	--------	--

Задание №16

Работа с файлом `country_vaccinations.csv`

Какое число указано в строке с индексом "3000" в столбце `daily_vaccinations_per_million`?

Источник данных - `country_vaccinations.csv`

Запишите ответ:

1)	Ответ:	
----	--------	--

Задание №17

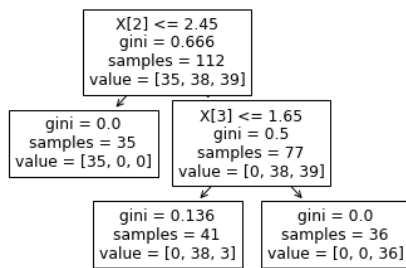
При обучении дерева решений на обучающей выборке, получена 100% точность классификатора, однако на тестовых классификатор показал точность около 50%. Что может быть причиной такого результата?

Запишите ответ:

1)	Ответ:	
----	--------	--

Задание №18

Перед вами дерево решений. Укажите самый высокий коэффициент загрязнения в листе (не во внутреннем узле)



Запишите ответ:

1)

Ответ:

Задание №19

Отберите данные за 5 февраля 2021 года. В какой стране количество привитых за день было максимально?

Источник данных - country_vaccinations.csv

Запишите ответ:

1)

Ответ:

Задание №20

Отсортируйте данные о вакцинации в Испании по полю daily_vaccinations. Используйте сортировку по убыванию. Укажите значение поля daily_vaccinations в третьей строке.

Источник данных - country_vaccinations.csv

Запишите ответ:

1)

Ответ:

Задание №21

Посчитайте в скольких наблюдениях в качестве вакцины использовался "Sputnik V" и "Pfizer".

Укажите оба числа через пробел. Сначала меньшее, потом большее.

Чтобы найти страны, в которых применялась вакцина, можно воспользоваться методом [Series.str.contains](#)

Источник данных - country_vaccinations.csv

Запишите ответ:

1)

Ответ:

Задание №22

Для ответа на этот вопрос вам понадобится датасет imdb из библиотеки keras

```
from keras.datasets import imdb
```

Определите какое из слов lucky или drama чаще встречается в датасете IMDb? УВ качестве ответа укажите абсолютную (по модулю) разницу между позициями этих слов в частотном словаре?

Запишите число:

1)

Ответ:

Задание №23

Сгруппируйте данные по дате наблюдения. Посчитайте сумму по полю daily_vaccinations. Укажите второе по величине значение.

Ответ - сумма, а не дата.

Источник данных - country_vaccinations.csv

Запишите ответ:

1)

Ответ:

Задание №24

Сгруппируйте данные по странам, для каждой страны вычислите сумму по полю "daily_vaccinations", отсортируйте полученные данные по убыванию. Укажите значение в третьей строке.

Ответ - одно число.

Источник данных - country_vaccinations.csv

Запишите ответ:

1)

Ответ:

Задание №25

Сгруппируйте данные по странам. Укажите страну, которая находится на втором месте по максимальному числу полностью вакцинированных (people_fully_vaccinated).

Источник данных - country_vaccinations.csv

Запишите ответ:

1)	Ответ:	
----	--------	--

Задание №26

Сколько людей было вакцинировано в России 10го февраля? Поле daily_vaccinations.

Источник данных - country_vaccinations.csv

Запишите ответ:

1)	Ответ:	
----	--------	--

Задание №27

Создайте случайные лес rf состоящий из15 деревьев. В ответе укажите строчку кода, отвечающую за его создание.

Запишите число:

1)	Ответ:	
----	--------	--

Задание №28

Имеется матрица ошибок:

Прогнозируемый	1	2	3	All
Фактический				
1	189	894	3028	4111
2	912	7195	30044	38151
3	4022	36797	223878	264697
All	5123	44886	256950	306959

Посчитайте показатель recall для класса с меткой 3.

Запишите число:

1)	Ответ:	
----	--------	--

Задание №29

Имеется матрица ошибок:

Прогнозируемый	1	2	3	All
Фактический				
1	189	894	3028	4111
2	912	7195	30044	38151
3	4022	36797	223878	264697
All	5123	44886	256950	306959

Посчитайте показатель precision для класса с меткой 2.

Запишите число:

1)	Ответ:	
----	--------	--

Задание №30

Загрузите встроенный набор данных digits.
from sklearn.datasets import load_digits
digs = load_digits()
Разделите данные на тестовый и тренировочный наборы в пропорции 75% и 25%.
Создайте дерево решений.
Выявите лучшие параметры дерева решений на следующих данных: максимальная глубина дерева от 2 до 10, минимальное количество выборок, необходимых создания листа (внешнего узла) от 4 до 6. В ответ запишите полученные показатели через пробел, сначала максимальную глубину, затем минимальное количество выборок, необходимых для создания узла.

Запишите ответ:

1)	Ответ:	
----	--------	--

Ответы:

#1 (1 б.)	Ответ = 'secret'
#2 (1 б.)	Ответ = 'loser'
#3 (1 б.)	Ответ = 6196
#4 (1 б.)	Ответ = 2075
#5 (1 б.)	Ответ = Falkland Islands

#6 (1 б.)	Ответ = 2034
#7 (1 б.)	Ответ = 143 3154 4024
#8 (1 б.)	Ответ = Python
#9 (1 б.)	Ответ = dt = DecissionTreeClassifier()
#10 (1 б.)	Ответ = 70598.0
#11 (1 б.)	Ответ = 2383.0
#12 (1 б.)	Ответ = 24.0
#13 (1 б.)	Ответ = 1.58
#14 (1 б.)	Ответ = 4041.0
#15 (1 б.)	Ответ = 29.16
#16 (1 б.)	Ответ = 2165.0
#17 (1 б.)	Ответ = Переобучение
#18 (1 б.)	Ответ = 0,136
#19 (1 б.)	Ответ = China
#20 (1 б.)	Ответ = 74783.0
#21 (1 б.)	Ответ = 391 3126
#22 (1 б.)	Ответ = 1589
#23 (1 б.)	Ответ = 6085565.0
#24 (1 б.)	Ответ = 17868335.0
#25 (1 б.)	Ответ = Israel
#26 (1 б.)	Ответ = 103571.0
#27 (1 б.)	Ответ = rf = RandomForestClassifier(n_estimators = 15)
#28 (1 б.)	Ответ = 0.845790
#29 (1 б.)	Ответ = 0.160295
#30 (1 б.)	Ответ = 9 4

Обучающийся владеет:

ОПК-10.3: Решает задачи в области профессиональной деятельности, используя перспективные методы машинного обучения	- навыками решения научно-технических задач в области своей профессиональной деятельности с использованием технологий анализа данных и машинного обучения.
--	--

Задание 31 1. Скачайте набор данных о поездах и транспортной сети Франции. https://www.kaggle.com/datasets/gatandubuc/public-transport-traffic-data-in-france. Прочтите данные из файла Regularities_by_liaisons_Trains_France.csv в датафрейм gb. Получите распределение числовых величин датафрейма gb используя базовые методы визуализации библиотеки pandas. Определите графически в каком месяце было больше всего отмененных поездов? Связаны ли пики данных с общим количеством запланированных поездов? Исследуйте графически корреляцию между числовыми признаками в наборе данных gb. Постройте тепловую карту. Задание 32 7. Скачайте данные из набора https://www.kaggle.com/code/ambaniverma/uk-traffic-accidents/data Используя средства визуализации, проведите исследование, есть ли параметры, которые наиболее сильно влияют на дорожные происшествия? В наборе не наблюдается сильных положительных корреляций, с погодой, местностью, возрастом водителя и т.д. Определите количество инцидентов в зависимости от зоны ограничения скорости. Задание 33 Зарегистрируйтесь в системе организации конкурсов по исследованию данных Kaggle. Скачайте набор данных о поездах и транспортной сети Франции. https://www.kaggle.com/datasets/gatandubuc/public-transport-traffic-data-in-france В разделе Details изучите информацию о файле Regularities_by_liaisons_Trains_France.csv. Ответьте на вопрос, в каком случае поезд считается прибывшим вовремя? Отсортируйте годы ведения статистики по общему количеству поездов, опоздавших более, чем на 15 минут. Выведите количество отмененных поездов по каждой станции отправления.	
--	--

2.3. Перечень вопросов для подготовки обучающихся к промежуточной аттестации

1. Машинное обучение.
2. Глубокое обучение.
3. Обучение с учителем и без учителя.
4. Основные понятия анализа данных. Данные, модель, обработка данных.
5. Методологические принципы анализа данных.
6. Цели, этапы, методы и техники анализа данных.
7. Фильтрация группировка и агрегация данных.
8. Визуализация данных.
9. Визуализация данных в Python.
10. Визуализация в Pandas.
11. Библиотеки Seaborn и Matplotlib
12. Решающие деревья: основные понятия.
13. Задача классификации.
14. Деревья решений. Основные параметры дерева.
15. Дерево решений в задаче регрессии.
16. Метод ближайших соседей.
17. Концепция случайного леса.
18. Случайная выборка тренировочных образцов.
19. Усреднение прогнозов.
20. Проблема переобучения.
21. Типы ошибок: true negative, false positive.
22. Метрики качества модели: Precision, Recall, F1 score.
23. Тестирование ROC AUC.
24. Нейронные сети.
25. Искусственные нейроны.
26. Функции активации нейрона.
27. Перцептрон.
28. Обучение нейрона.
29. Градиентный спуск.
30. Дробление шага при градиентном спуске.
31. Стохастический градиентный спуск.
32. Метод наискорейшего спуска.
33. Нейронная сеть.
34. Однослойная модель.
35. Сверточные нейронные сети.
36. Свёрточный слой, слой подвыборки, полносвязный слой.
37. Целевая функция.
38. Задачи распознавания образов.
39. Классификация объектов.
40. Поиск изображения по образцу.

3. Методические материалы, определяющие процедуру и критерии оценивания сформированности компетенций при проведении промежуточной аттестации

Критерии формирования оценок по ответам на вопросы, выполнению тестовых заданий

- оценка **«отлично»** выставляется обучающемуся, если количество правильных ответов на вопросы составляет 100 – 90% от общего объема заданных вопросов;
- оценка **«хорошо»** выставляется обучающемуся, если количество правильных ответов на вопросы – 89 – 76% от общего объема заданных вопросов;
- оценка **«удовлетворительно»** выставляется обучающемуся, если количество правильных ответов на тестовые вопросы – 75–60 % от общего объема заданных вопросов;
- оценка **«неудовлетворительно»** выставляется обучающемуся, если количество правильных ответов – менее 60% от общего объема заданных вопросов.

Критерии формирования оценок по результатам выполнения заданий

«Отлично» – ставится за работу, выполненную полностью без ошибок и недочетов.

«Хорошо» – ставится за работу, выполненную полностью, но при наличии в ней не более одной негрубой ошибки и одного недочета, не более трех недочетов.

«Удовлетворительно» – ставится за работу, если обучающийся правильно выполнил не менее 2/3 всей работы или допустил не более одной грубой ошибки и двух недочетов, не более одной грубой и одной негрубой ошибки, не более трех негрубых ошибок, одной негрубой ошибки и двух недочетов.

«Неудовлетворительно» – ставится за работу, если число ошибок и недочетов превысило норму для оценки «удовлетворительно» или правильно выполнено менее 2/3 всей работы.

Виды ошибок:

- *грубые ошибки: незнание основных понятий, правил, норм; незнание приемов решения задач; ошибки, показывающие неправильное понимание условия предложенного задания.*
- *негрубые ошибки: неточности формулировок, определений; нерациональный выбор хода решения.*
- *недочеты: нерациональные приемы выполнения задания; отдельные погрешности в формулировке выводов; небрежное выполнение задания.*

Критерии формирования оценок по зачету

«Зачтено» – студент демонстрирует знание основных разделов программы изучаемого курса: его базовых понятий и фундаментальных проблем; приобрел необходимые умения и навыки, освоил вопросы практического применения полученных знаний, не допустил фактических ошибок при ответе, достаточно последовательно и логично излагает теоретический материал, допуская лишь незначительные нарушения последовательности изложения и некоторые неточности.

«Незачтено» – выставляется в том случае, когда студент демонстрирует фрагментарные знания основных разделов программы изучаемого курса: его базовых понятий и фундаментальных проблем. У экзаменуемого слабо выражена способность к самостоятельному аналитическому мышлению, имеются затруднения в изложении материала, отсутствуют необходимые умения и навыки, допущены грубые ошибки и незнание терминологии, отказ отвечать на дополнительные вопросы, знание которых необходимо для получения положительной оценки.